RÉSUMÉ DE THÈSE

Etude de l'impact de l'échantillonnage des espèces et de la longueur des séquences sur la robustesse des phylogénies moléculaires. Implications sur la phylogénie des Téléostéens, par Guillaume **Lecointre**, Laboratoire d'Ichtyologie générale et appliquée, Muséum national d'Histoire naturelle, 43 rue Cuvier, 75231 Paris cedex 05, FRANCE.

Thèse de Doctorat ès Sciences, Univ. Paris VII, 1993, 445 pp., 17 tab., 97 fig., 358 réf.

La phylogénie et la classification des Poissons, **taxon paraphylétique** de Vertébrés, furent bouleversées par l'introduction précoce de la cladistique en Ichtyologie. On rappelle que la classification phylogénétique (au sens de Hennig) est l'expression moderne de la pensée darwinienne. Si l'Ichtyologie systématique est depuis 25 ans l'héritière de Darwin et de Hennig en matière de classification, elle ne reçut en revanche que peu d'apports de la Systématique moléculaire, discipline trop jeune s'étant focalisée sur un nombre limité de problématiques. Les tendances extrêmement réductionnistes de la systématique moléculaire sont soulignées: celle-ci n'apporte des résultats phylogénétiques fiables que si certaines précautions ont été prises en matière d'échantillonnage taxinomique.

La phylogénie des Poissons est étudiée à partir d'une base de données de séquences d'ARN ribosomique 28S, à laquelle de nouvelles séquences ont été ajoutées. Sa richesse taxinomique a permis d'étudier précisément l'ampleur et la répartition des variations des valeurs de **bootstrap** causées par le changement ou l'ajout d'espèces dans un arbre. Les arbres phylogénétiques à 4 espèces ne sont pas fiables car ils présentent des parentés contradictoires robustes en fonction des espèces choisies pour représenter les mêmes taxa. Une procédure originale d'échantillonnage exhaustif d'une espèce par groupe prédéfini montre l'impact de chaque espèce sur la robustesse de chaque noeud. Elle permet d'établir que l'impact de l'échantillonnage d'un groupe donné est trés localisé dans l'arbre, limité à son point d'insertion et aux noeuds immédiatement voisins. Ces conclusions sont généralisables à des bases de données très diverses.

Une étude détaillée de la qualité et de la répartition de l'**homoplasie** dans les séquences permet d'établir que l'homoplasie effective est, pour un site donné, régionalisée dans l'arbre et que cette seconde régionalisation pourrait fort bien expliquer la première. Cette régionalisation de l'homoplasie révèle la nécessité de mouvelles méthodes de mesure de l'homoplasie et de pondération, ne prenant plus en compte le site globalement mais les évènements locaux qu'il contient.

L'étude de l'évolution de la robustesse de chaque noeud en fonction de la longueur des séquences a notamment permis la création d'un outil d'inférence du nombre de sites informatifs requis pour la résolution d'un noeud donné.

Les principales conséquences phylogénétiques de ce travail sont au nombre de trois. Premièrement, les séquences d'ARNr 28S semblent attester, avec d'autres données, l'apparition quasi-simultanée des trois grandes lignées de Gnathostomes actuels (Chondrichthyens, Actinoptérygiens, Sarcoptérygiens) il y a 420 millions d'années. Deuxièmement, les Elopomorphes sont peut-être le groupe-frère des Ostéoglossomorphes, mais cette parenté n'est attestée ni par les données paléontologiques ni par les données morpho-anatomiques. En revanche, il se peut fort bien que les Ostariophyses soient le groupe-frère des Clupéomorphes, même si cette parenté sur le plan moléculaire risque d'avoir été obtenue par artefact d'attraction des branches longues. En effet, des caractères morpho-anatomiques plaident pour un tel regroupement (notamment la fusion de l'hypural 2 au centre ural 1 et la fusion extrascapulaire-pariétal), tandis que ceux qui le contredisent ne sont pas fiables. Le clade des Eutéléostéens demande à être redéfini.

Définitions. - **Bootstrap**: outil statistique de mesure de la robustesse des arbres phylogénétiques, fondé sur un tirage avec remise des caractères. - **Homoplasie**: toute similitude qui n'est pas acquise par ascendance commune. - **Taxon paraphylétique**: taxon comprenant un ancêtre et seulement une partie de ses descendants connus. S'oppose au taxon monophylétique, qui comprend un ancêtre et tous ses descendants connus.

Publications correspondantes. - *Mol. Phylogen. Evol.*, 1993: 2(1): 31-51. - *Mol. Phylogen. Evol.*, 1993: 2(3): 205-224. - *Mol. Phylogen. Evol.*, 1994: in press.

Summary. - Impacts of species sampling and sequence length on robustness of molecular phylogenies. Consequences for teleostean phylogeny.

The phylogeny and classification of fishes, a **paraphyletic taxon** of Vertebrates, were strongly affected by the early practice of cladistics in Ichthyology. Phylogenetic classification (in Hennig's sense) is a modern expression of the darwinian tradition. It is stressed that systematic Ichthyology has been for 25 years within the darwinian-hennigian tradition. The impact of molecular systematics on systematic Ichthyology is, for the moment, limited to the resolution of few problems. It is stressed that the pitfall of reductionism of species samples is a weakness of molecular systematics and it is shown that its results are reliable only when they are based on a sufficient taxinomic sample.

The phylogeny of fishes is studied through a data base of 28S rRNA, to which new sequences have been added. The richness of this taxinomic sample allowed the precise study of the amplitude and distribution of the variations in **bootstrap** proportions provoked by variations in species sampling. It is shown that 4-species trees are not reliable because they can strongly support contradictory relationships depending on the species sample chosen to represent the same taxa. An original procedure of exhaustive sampling of a single species per presumed monophyletic group has shown precisely the impact of each species on each node, and allowed to show that the impact of species sampling of a given group is strongly localized, limited to its branching node in the tree and its two neighbouring nodes. These conclusions are generalizable to various other molecular data sets.

A study of the quality and distribution of **homoplasy** within the sequences allowed to show that effective homoplasy, for a given site, is localized in a sub-part of the tree, and that such a pattern could explain the localized impact of species sampling. The distribution of homoplasy reveals the need of new methods of homoplasy measurement and character weighting, not to be applied on entire sites but on local evolutionary events.

The study of evolution of the robustness with the increasing sequence length allowed to propose a tool to infer the number of informative sites required for the resolution of a given node.

The main phylogenetic consequences of this study are threefold. First, 28S rRNA sequences tends to indicate (with other molecular and paleontological data) that the three main gnathostome lineages (Chondrichthyans, Actinopterygians, Sarcopterygians) appeared simultaneouly 420 millions years ago. Second, Elopomorphs and Osteoglossomorphs may be sister-groups, although this relationship is confirmed neither by paleontological data nor morphological data. Third, Ostariophysans-Clupeomorphs sister-group relationships should be seriously considered. The molecular grouping could have been obtained from a long-branch-attraction artifact, but there are morphological features that could support this relationship (at least the hypural 2-ural 1 fusion and the parietal-extrascapular fusion), and those which contradict it are not valid. The euteleostean clade needs to be redefined.

Key-words. - Bootstrap, Character sampling, Fishes, Molecular phylogeny, Teleostean phylogeny.